

AD-A088 801

DYNAMICS RESEARCH CORP WILMINGTON MASS

F/G 5 9

AIR FORCE PERSONNEL AVAILABILITY ANALYSIS: APPLICATION TECHNIQUE--PTC(U)

AUG 80 J C GOCLOWSKI, S E PESKOE, A J LOFASO F33615-77-C-0032

UNCLASSIFIED

AFHRL-TR-79-67

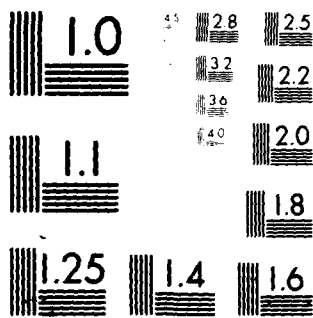
NL

2

END
DATE
-10-10
/11C

Cont

8880



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AIR FORCE



HUMAN RESOURCES

AD A088801

DC FILE COPY

LEVEL

②
SC

**AIR FORCE PERSONNEL AVAILABILITY ANALYSIS:
APPLICATION TECHNIQUES OF THE PERSONNEL
AVAILABILITY MODEL (PAM)**

By

**John C. Goclowski
Stuart E. Peskoe
A. J. LoFaso**

**Dynamics Research Corporation
60 Concord Street
Wilmington, Massachusetts 01887**

H. Anthony Baran

**LOGISTICS AND TECHNICAL TRAINING
Wright-Patterson Air Force Base, Ohio 45433**

August 1980

Final Report

Approved for public release; distribution unlimited.

DTIC
SELECTED
AUG 28 1980

LABORATORY

**AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235**

80 9 4 035

NOTICE

When U.S. Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This final report was submitted by Dynamics Research Corporation, 60 Concord Street, Wilmington, Massachusetts 01887, under Contract F33615-77-C-0032, Project 1959, with the Logistics and Technical Training Division, Air Force Human Resources Laboratory (AFSC), Wright-Patterson Air Force Base, Ohio 45433. Mr. H. Anthony Baran was the Contract Manager for the Laboratory.

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

ROSS L. MORGAN, Technical Director
Logistics and Technical Training

RONALD W. TERRY, Colonel, USAF
Commander

SUBJECT TO EXPORT CONTROL LAWS

This document contains information for manufacturing or using munitions of war. Export of the information contained herein, or release to foreign nationals within the United States, without first obtaining an export license, is a violation of the International Traffic in Arms Regulations. Such violation is subject to a penalty of up to 2 years imprisonment and a fine of \$100,000 under 22 U.S.C. 2778.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

19 REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFHRL-TR-79-67	2. GOVT ACCESSION NO. AD-A088	3. RECIPIENT'S CATALOG NUMBER 802
4. TITLE (and Subtitle) AIR FORCE PERSONNEL AVAILABILITY ANALYSIS: APPLICATION TECHNIQUES OF THE PERSONNEL AVAILABILITY MODEL (PAM).		5. TYPE OF REPORT & PERIOD COVERED Final / rept. g
7. AUTHOR(s) John C. Goculowski Stuart E. Peskoe A. J. LoFaso H. Anthony Baran		8. CONTRACT OR GRANT NUMBER(s) F33615-77-C-0032
9. PERFORMING ORGANIZATION NAME AND ADDRESS Dynamics Research Corporation 60 Concord Street Wilmington, Massachusetts 01887		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 63751F 19590003
11. CONTROLLING OFFICE NAME AND ADDRESS HQ Air Force Human Resources Laboratory (AFSC) Brooks Air Force Base, Texas 78235		12. REPORT DATE August 1980
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Logistics and Technical Training Division Air Force Human Resources Laboratory Wright-Patterson Air Force Base, Ohio 45433		13. NUMBER OF PAGES 54
		15. SECURITY CLASS. (of this report) Unclassified
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) human resource availability human resource requirements maintenance personnel Markov model personnel availability model personnel availability weapon system maintenance		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This is the second of three technical reports describing a methodology for projecting the future availability of Air Force personnel and analyzing the potential impacts of personnel policy changes. Developed within the Air Force Human Resources Laboratory's Project 1959, "Advanced System for the Human Resources Support of Weapon System Development," the methodology will provide the Air Force with an increased capability for considering the human resources requirements of weapon systems in terms of the future availability of personnel for maintaining and operating the systems and the factors which determine their availability. Technical report AFHRL-TR-79-66 documents the development of the methodology for a computerized Personnel Availability Model (PAM) and its associated data bank. This report (AFHRL-TR-79-67) describes		

DD FORM 1 JAN 73 1473

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

388902

Handwritten signature/initials.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

Item 20 (Continued)

application techniques of the PAM. Technical report AFHRL-TR-79-68 provides a program description which includes an applied PAM analysis of Air Force personnel.

The PAM represents career transition activity within the Air Force according to a series of Markov processes. Each process depicts a subpopulation of airmen with states defined by year of service (YOS) and paygrade. The PAM uses a mathematical Markov model to project future personnel availability. It assumes that, in respect to career transition, the population of technical personnel is homogeneous within the Air Force. This report describes the results of the effort undertaken to check the validity of the Markov model approach to Air Force personnel availability analysis and also that of the homogeneity assumption.

Two procedures were developed as part of this effort. The first involves the identification of personnel attributes which may impact career transition rates of the Air Force technical personnel. The second procedure involves accurate projections of the described Air Force technical personnel subpopulations if impacting personnel attributes are identified.

No impacting attributes were found for the 13 technical Air Force Specialty Codes investigated. Thus, the validity of the Markov model and the homogeneity of technical personnel are accepted with regard to career transition. In addition, two byproducts of this validation effort tend to widen the range of PAM application. The first byproduct indicated that the procedures developed in this effort also may be applied to nontechnical Air Force personnel. Thus, the PAM also may be used to project the future availability of all Air Force personnel. The second byproduct stems from the statistical techniques used in this effort. It indicated that one of the statistical techniques is useful for modifying transition probabilities based on expected or proposed changes in Air Force personnel policy.

Accession For	
NTIS GIL&I	<input checked="checked" type="checkbox"/>
DDC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distributor/ _____	
Availability _____	
Dist	Availability Special
A	

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

SUMMARY

BACKGROUND

To assure that future weapon systems operational requirements are met, the following needs must be fulfilled:

1. An estimation of the personnel requirements which will arise from the implementation of future weapon systems
2. Identification of possible difficulties in fulfilling future requirements
3. Development of techniques capable of defining and identifying corrective actions which can be taken to avert, or at least diminish, the effect of future personnel shortages on weapon operation

This technical report describes the results of an effort undertaken as an attempt to satisfy these needs. Within that effort, the Personnel Availability Model (PAM) was developed to project the future availability of Air Force personnel on the basis of historical data recorded in the Uniform Airman Record (UAR). This report describes techniques identified in that effort with the ability to improve the projection capability of the PAM and extend its productivity in the analysis of Air Force personnel availability.

APPROACH

The PAM uses a Markov process to represent the career transition activity of Air Force personnel. The Markov process also serves as the basis for projection of transition activity to predict the future status of personnel. Use of this approach is based on an assumption that the population under examination is homogeneous in terms of its career transition behavior. To verify this assertion, an effort was undertaken to examine the PAM data base sample of Air Force personnel and determine (a) whether that population violates the homogeneity assumption necessary for its application to the analysis of Air Force personnel availability, and (b) whether attributes of Air Force personnel could be identified to define subpopulations which do not violate the homogeneity assumption. In short, the objective of the effort described in this report was to check the validity of the Markov model and, thus, the homogeneity assumption for technical personnel.

To achieve this objective, two procedures were developed.

Procedure One Applies statistical techniques to the PAM data base. It is designed as a search procedure to identify personnel attributes which might impact personnel transition (impacting attributes).

Procedure Two Makes accurate availability projections if impacting personnel attributes are identified. It involves the development of availability projections based on homogeneous subpopulations, as defined by their impacting attributes.

Note that accurate projections can be made only if the population in question is homogeneous as tested by the two procedures.

RESULTS

Impacting attributes recorded in the UAR, which would indicate a need for identifying homogeneous subpopulations, could not be found in the existing PAM data base for technical personnel. Therefore, the accuracy of the Markov model formulation for technical personnel and the assumption of homogeneity was supported.

Two statistical techniques are used in the first of the two procedures. These techniques are capable of:

1. Extending the range of the PAM application; checking the validity of the Markov and homogeneity assumptions; and, directly applying the PAM to all Air Force personnel (technical and nontechnical).
2. Modifying the PAM transition probabilities so that the impact of any expected or proposed changes in Air Force personnel policies can be evaluated through PAM projections.

PREFACE

This is the second of three technical reports which describe the Personnel Availability Model (PAM), its related application methodology, and application techniques. The model is designed for use in projecting and analyzing the future availability of Air Force personnel. Work was performed under USAF Contract No. F33615-77-C-0032.

Technical report AFHRL-TR-79-66 presents the development and functions of the PAM and its related data bank. This report (AFHRL-TR-79-67) describes application techniques of the model. Report AFHRL-TR-79-68 provides a program description which includes a PAM analysis of Air Force personnel.

Work was directed by the Advanced Systems Division, Air Force Human Resources Laboratory, Wright-Patterson Air Force Base, Ohio. It is documented under Work Unit 19590003 of AFHRL Project 1959, "Advanced System for Human Resources Support of Weapon Systems Development." Dr. William B. Askren was the Project Scientist. Mr. H. Anthony Baran was the Work Unit Scientist and Air Force Contract Manager. Mr. John Goclowski was the Contractor Program Manager.

Work Unit 19590003, "Air Force Personnel Availability Analysis," was undertaken to provide the Air Force with improved tools and techniques for anticipating the future impact on its personnel force structure resulting from the human resource requirements of new weapon systems.

Appreciation is extended to Dr. Gordon A. Eckstrand and Dr. Ross L. Morgan of the Advanced Systems Division for their guidance in constructing the modeling system described in this report, and to Dr. Robert A. Bottenberg of the Computation Sciences Division of the Air Force Human Resources Laboratory for his help in obtaining the personnel data essential to its operation.

TABLE OF CONTENTS

	Page
INTRODUCTION	7
PROCEDURE ONE: IDENTIFYING PERSONNEL ATTRIBUTES WHICH IMPACT TRANSITION	9
The Technique of Categorical Analysis	9
The Technique of Dependency Analysis	19
Summary of Procedure One	20
PROCEDURE TWO: MAKING ACCURATE AVAILABILITY PROJECTIONS	21
DATA PROCESSING	23
MODIFYING TRANSITION PROBABILITIES	29
EXTENDING THE RANGE OF THE PAM	31
CONCLUSION	33
REFERENCES	35
APPENDIX A - Logit Analysis	37
APPENDIX B - Discriminant Analysis	41
REFERENCES TO APPENDIX A AND B	43
APPENDIX C - Confidence Limits	45

LIST OF FIGURES

		Page
1	The PAM methodology Reflecting Procedures One and Two	10
2	Frequency Distribution of AFQT for the 431xx AFSC First-Term Enlistees in 1975	12
3	Loss Frequencies by AFQT	13
4	Losses by AFQT (Hypothetical)	14
5	Dichotomies	15
6	Polytomies	16
7	Discrete (or Pseudo-continuous)	17

LIST OF TABLES

		Page
1	Variable Types	18
2	Data File	24
CI	Confidence Limits	48

INTRODUCTION

The effort described in this technical report was conducted to develop tools and techniques that would extend the Air Force capability to analyze the availability of its personnel and estimate their availability in future time-frames. The purpose was to fulfill a need to compare the manpower requirements of new weapon systems with projections of manpower availability at a time when personnel would be required to support the operation of those systems. The goal of such a comparison is to reduce or avoid the incidence of personnel shortages through system development planning while considering the implications of both design and support personnel availability.

In developing the tools and techniques, the career transition activity of Air Force personnel was modeled in terms of a Markov process. The result was the Personnel Availability Model (PAM) described in AFHRL-TR-79-66. Use of a Markov formulation was predicated on several assumptions, which also are described in that report with respect to the prediction of career progression behavior in Air Force personnel. The assumptions include the existence of relationships between career transition activity of Air Force personnel and descriptors of those personnel (personnel attributes) as recorded in the Uniform Airman Record (UAR).

The focus of PAM effort was to determine if patterns of similar career transition activity could be predicted on the basis of the personnel attributes of an airman population, and if identifiable subpopulations could be determined to behave in a similar manner. Techniques were identified and applied to determine (a) the statistical significance of personnel attributes as determinants of career transition behavior; and (b) the relevance of such determinants, if found, to the solution of real world personnel availability prediction problems within the Air Force. The results include:

1. Confirmation of the applicability of the Markov process to PAM requirements
2. Extension of the PAM's capability to address the prediction of future availability for nontechnical as well as technical, Air Force Specialty Codes (AFSCs)
3. Identification and development of two procedures to increase the utility of the PAM

The two resulting procedures were defined to determine whether relationships existed between personnel career transition activity and personnel attributes which might either detract from the propriety of a Markov process for use in personnel availability analysis or be used to increase the accuracy of PAM predictions or extend its applicability. It entailed a search for statistical analysis techniques which could be applied to the PAM data base to identify homogeneous behavior

on the part of personnel who could be described in terms of a single attribute or a combination of attributes. It should be noted that the sole objective in isolating such "driving attributes" was to establish a basis for statistical inference in the prediction of transition activity, and not to ascribe causality or quantify those relationships which might be found to exist.

Procedure One can be used to determine whether personnel attributes significantly affect, or can be used to predict, the career transition activity of Air Force maintenance personnel. Procedure Two can be used to quantify the aggregative effects of the combinations of personnel attributes found to be related to career transition activity. It can also be used to determine what changes should be made in the transition probability matrix used by the PAM to project the future availability of personnel on a year-by-year basis.

The purpose of this report is to provide a greater understanding of relationships between career transition behavior of personnel and their characteristics (driving attributes). In terms of Procedures One and Two, this report intended to further examine the merits of the assumptions made in the development of the PAM, and identify techniques to increase the accuracy of personnel availability predictions and extend the limits of applicability.

PROCEDURE ONE: RECOGNIZING TRANSITION IMPACTING ATTRIBUTES

Two techniques are used in Procedure One to recognize transition impacting attributes. They are categorical analysis and dependency analysis, as described in this section. (Because general purpose statistical packages already exist in the Air Force inventory of computer programs, the techniques used in this study are not included in the PAM computer program.)

Figure 1 presents the PAM projection process incorporating Procedures One and Two. Note that Procedure One is a part of a branch which may be required if lack of homogeneity is suspected. This branch is made following an inspection of the abbreviated UARs for those personnel involved in the projection.

THE TECHNIQUE OF CATEGORICAL ANALYSIS

The purpose of categorical analysis is to identify homogeneous subpopulations with significantly different transition rates. Since such groups are defined by the personnel attributes they share in common, the main task of the analysis is a systematic search for driving attributes. Thus, the goal of categorical analysis is to determine the combinations of driving attributes necessary to make divisions (in terms of transition rates) among the population.

The categorical analysis technique was designed for the following hypothetical problem type (the histograms mentioned below are described later in this section).

It has been found that 10 percent of the males in AFSC 328x0 left the Air Force in 1975. The distribution of these losses across each of the remaining attributes is examined to find which, if any, show nonuniform loss rates. Histograms of loss frequency versus other attributes are suitable displays for this purpose. Furthermore, among the several displays, the one for losses versus school (years of schooling beyond eighth grade) shows that 90 percent of the total losses were college graduates. Another histogram of 328x0 males by school indicates that 20 percent of the 328x0 males are college graduates. These indicate a predominantly noncollegiate population with a preponderance of college graduates among the losses. It follows that school is an attribute that distinguishes between two similar subpopulations with disparate loss rates; therefore, school is a driving attribute.

Other transitions and subpopulations may be investigated in much the same manner to identify the driving attributes for other state/transition combinations. Thus, the analysis represents a systematic comparison of transition rates as frequency distributions over individual

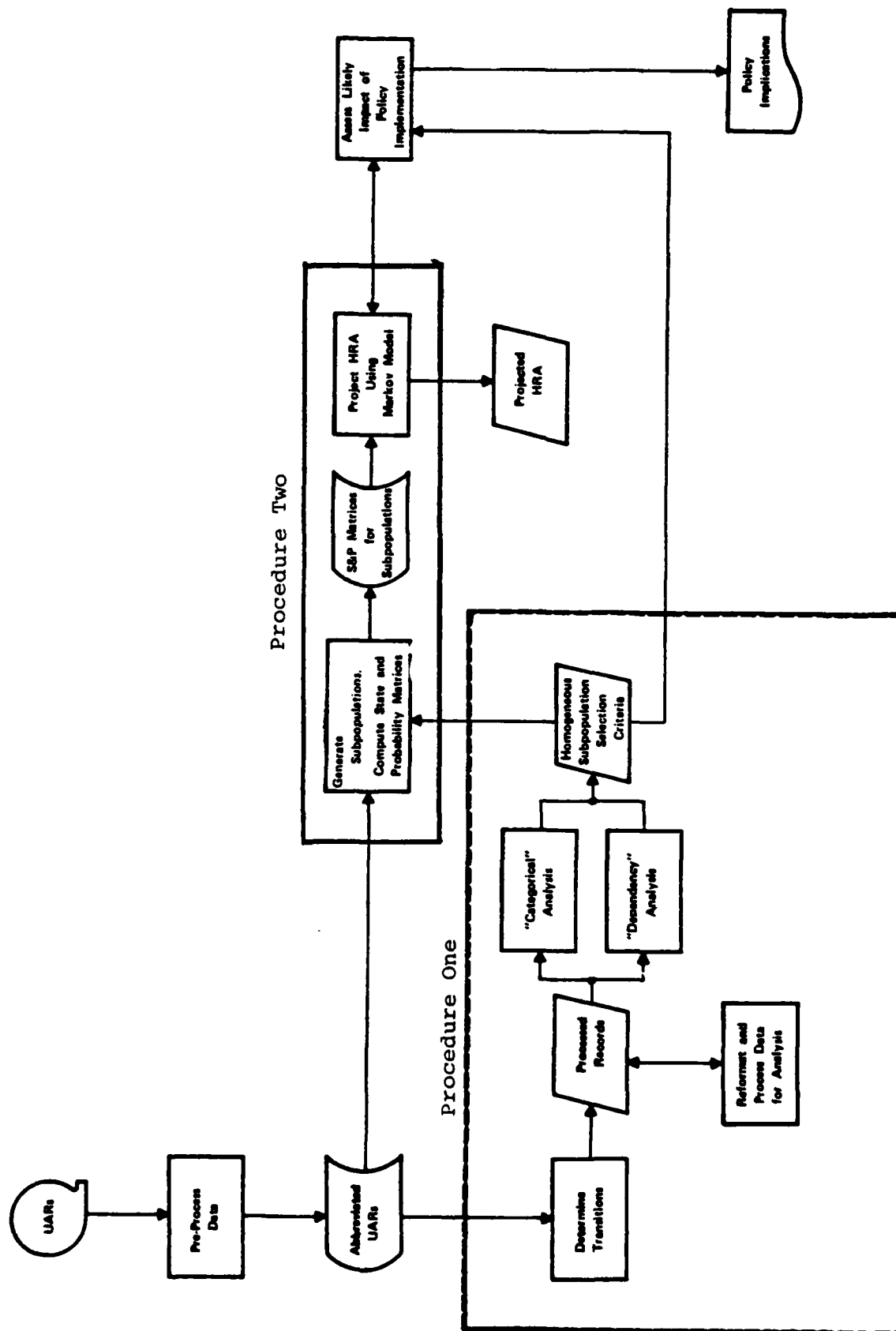


Figure 1 — The PAM methodology reflecting procedures one and two.

attributes, a qualitative discriminant analysis where the presence or absence (high or low values) of specific attributes is indicative of sub-population membership.

The statistical portion of categorical analysis is not complex. The major tools used in the analysis of categorical data are histograms generated by the Statistical Analysis Subsystem (SASS). Figures 2, 3, and 4 provide examples of histograms. The histogram in Figure 2 is a plot of the frequency distribution of the Armed Forces Qualification Test (AFQT) composite of the Armed Services Vocational Aptitude Battery (ASVAB), for the 431xx AFSC first-term enlistees in 1985. The ordinate is the frequency of occurrence and the abscissa is the possible values taken by the variable AFQT. Figure 3 shows the distribution of losses for that same group of airmen. In itself, Figure 3 offers little insight into the effect of AFQT on the loss rate. However, Figure 4 hypothetically shows the distribution of losses over AFQT, indicating that AFQT is indeed a driving attribute.

Categorical analysis involves histograms and simple t-statistic tests to determine the significance of differences between the means of the distributions compared. The key to the analysis is comparison. If the distributions of transitions and populations over the same variable are similar, that variable is not a driving attribute; if the distributions show disparate transition rates, the variable is likely to be a driving attribute. The product of this analysis is a sequenced list of driving attributes which contain all the information necessary to cross-classify the airmen into homogeneous subpopulations. However, there is a subjective decision to be made: How different must the distributions be to indicate discrimination?

The terms "similar" and "disparate" are vague. Therefore, criteria must be set for determining when distributions are different. It was decided that a significant difference in the means of two distributions (using a t-test at the .05 level) is sufficient to indicate that the population is heterogeneous in relation to transition rates and must be divided into subpopulations to project personnel availability.

Another problem is how to make the division. For dichotomous or polytomous variables (see Table 1) the divisions are obvious--the population is split along the dichotomies and relevant polytomies. The discrete variables must be divided into intervals (for example, years of service (YOS) may be broken into first enlistment, second enlistment, and career personnel if results indicate that this is reasonable). There should be as many intervals as there are subpopulations discernible within the given population. The boundary locations may be natural, as in YOS above; regular, as in quartiles; or artificial, when the boundaries are well suited to the distribution exhibited, but there is no natural break. Figures 5 and 6 illustrate how dichotomies and polytomies may be categorized. Figure 7 gives two examples of interval splitting of discrete

01/26/78

AFSC 431XX 1975
1-5 YOS
FREQUENCIES -- AFQT

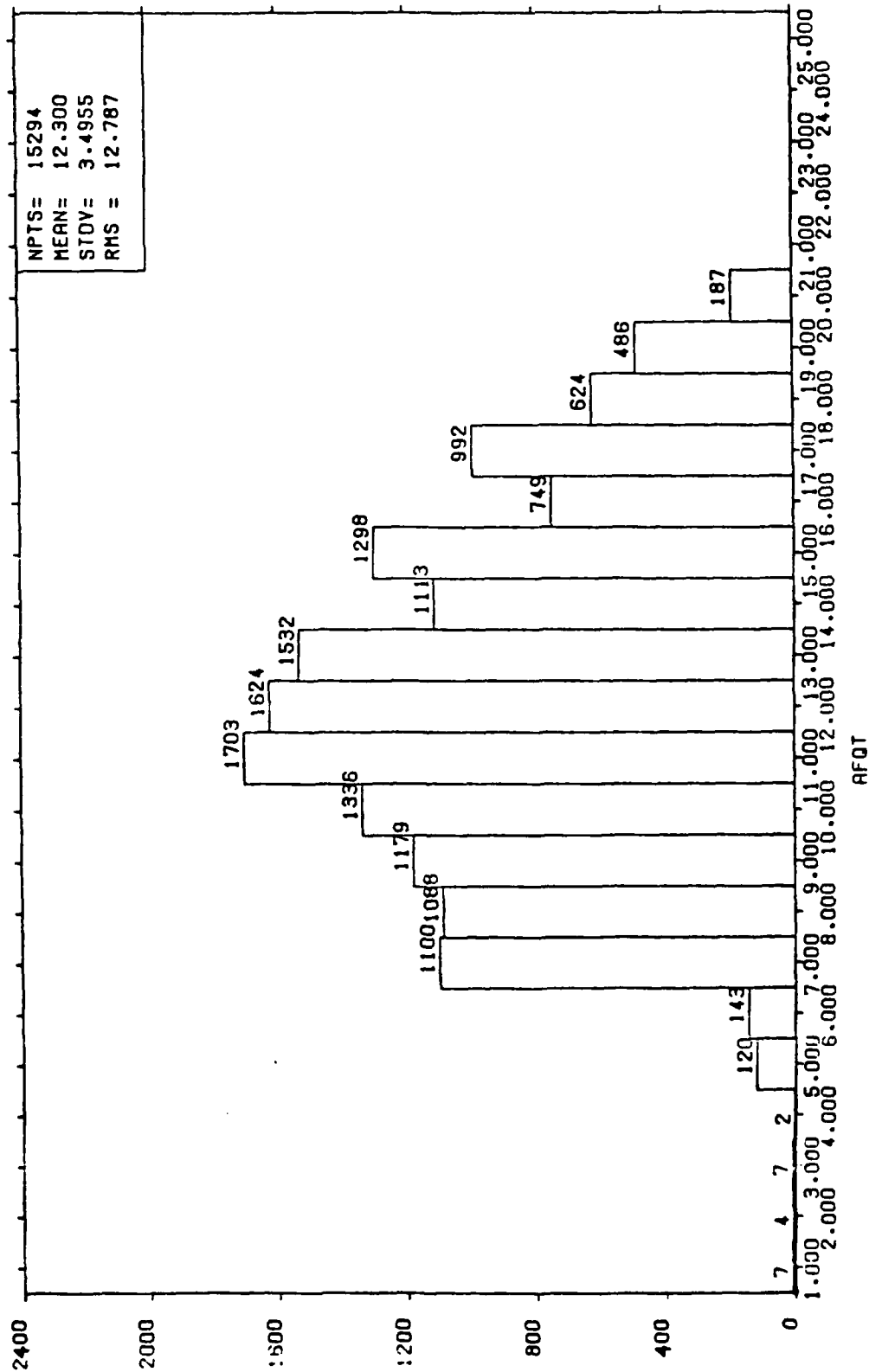


Figure 2 -- Frequency distribution of AFQT scores for the 431xx AFSC first-term enlistees in 1975.

01/26/78

AFSC 431XX 1975
1-5 YOS
LOSS FREQUENCIES BY AFQT

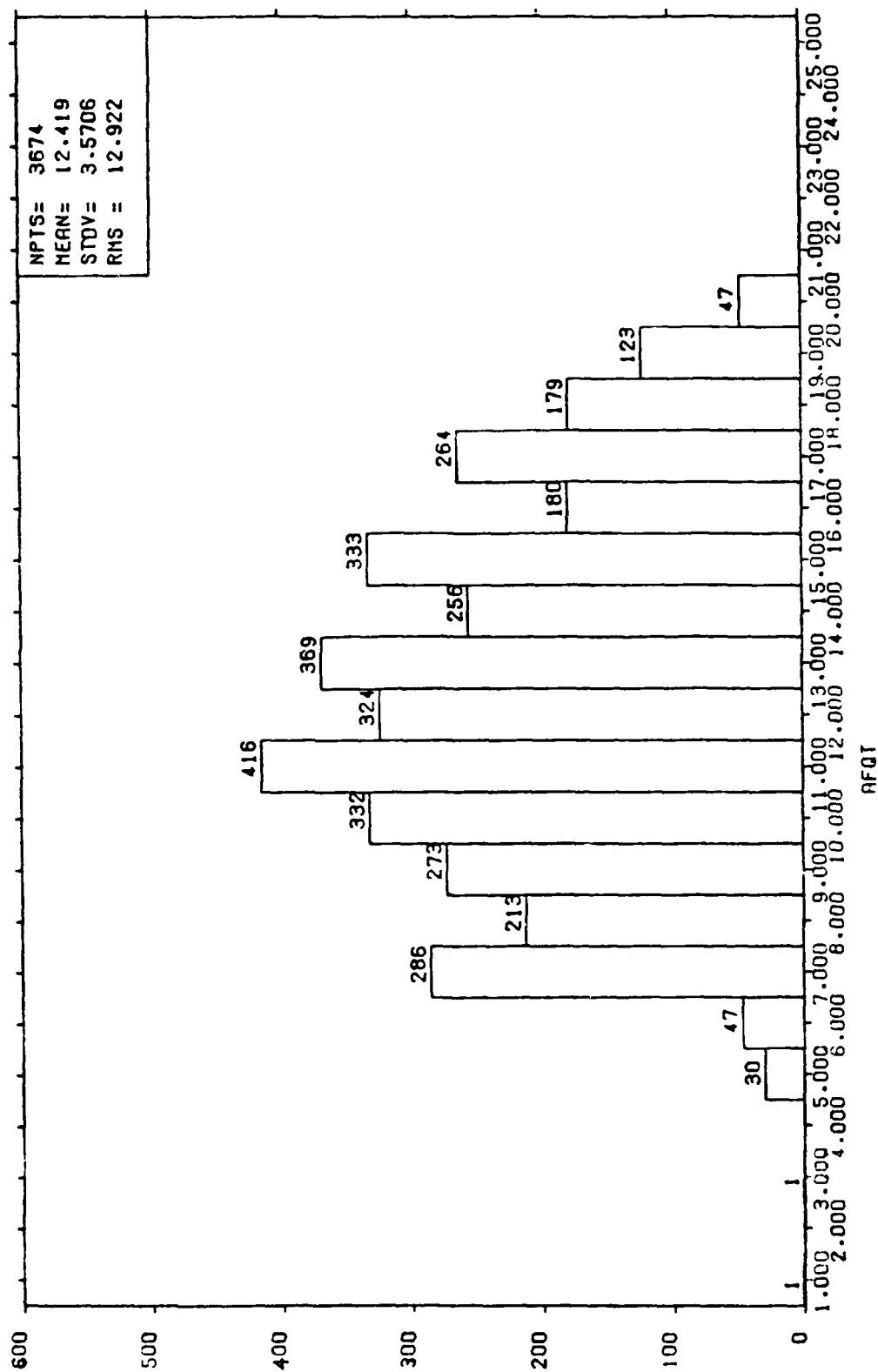


Figure 3 - Loss frequencies by AFQT.

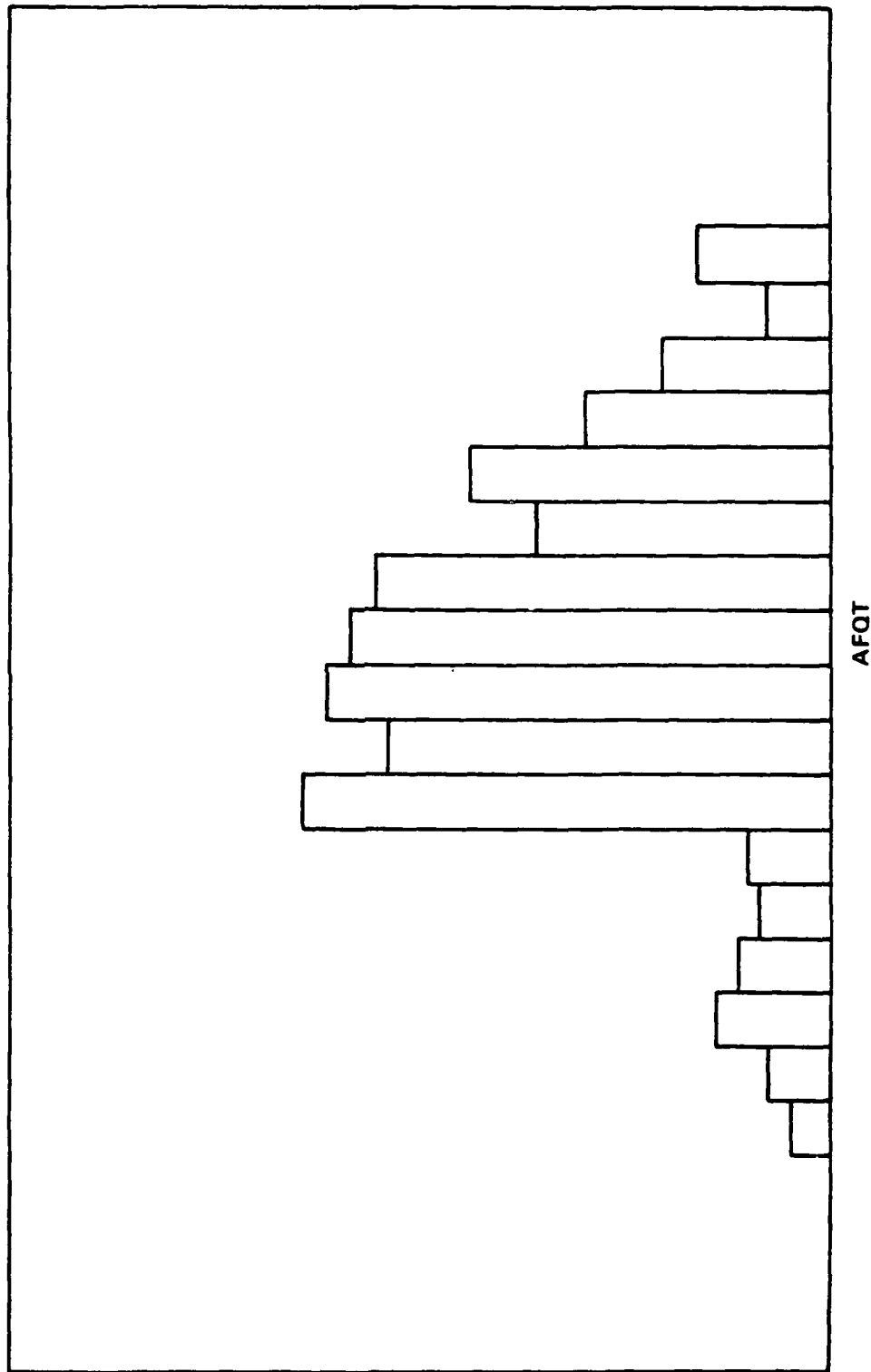
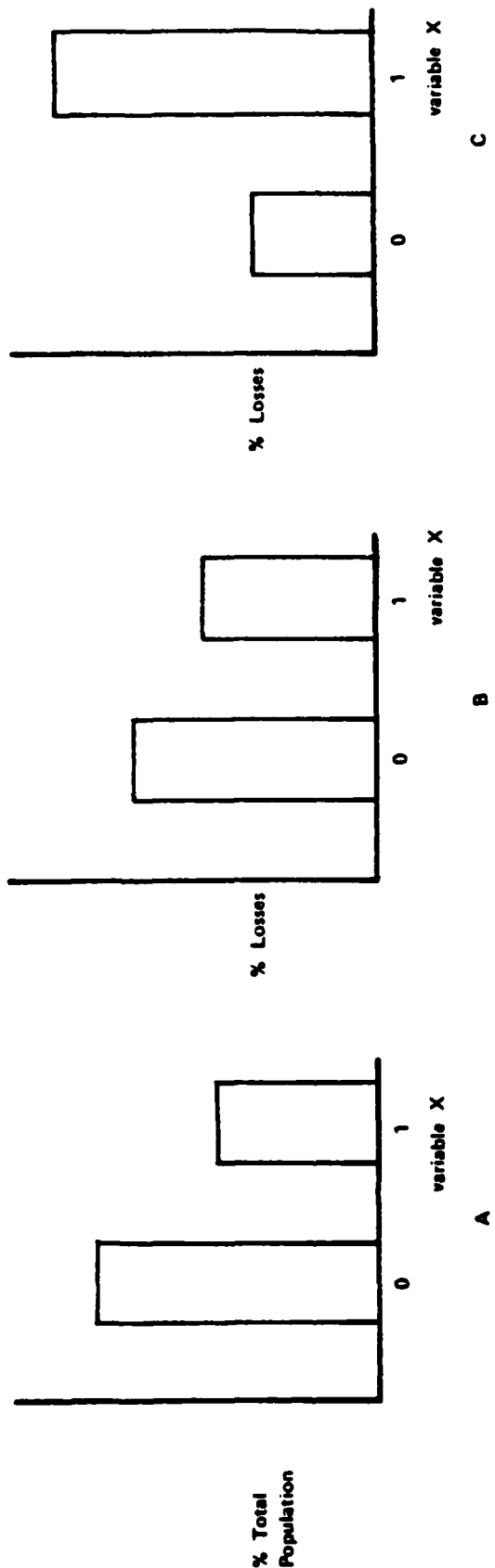


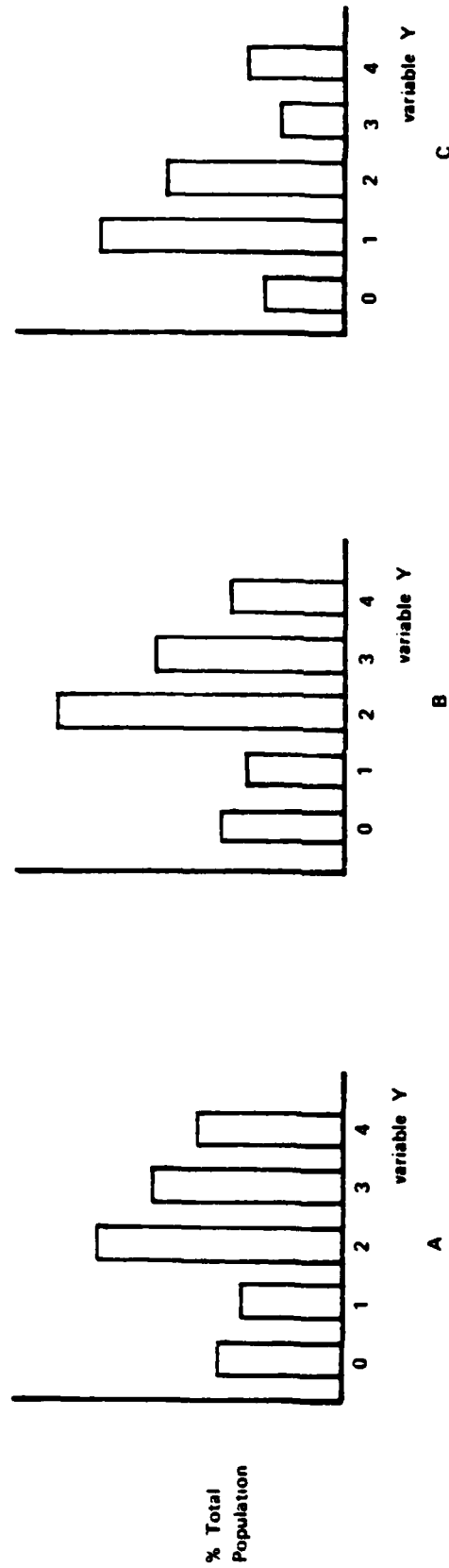
Figure 4 - Losses by AFQT (Hypothetical).

OF LOSSES



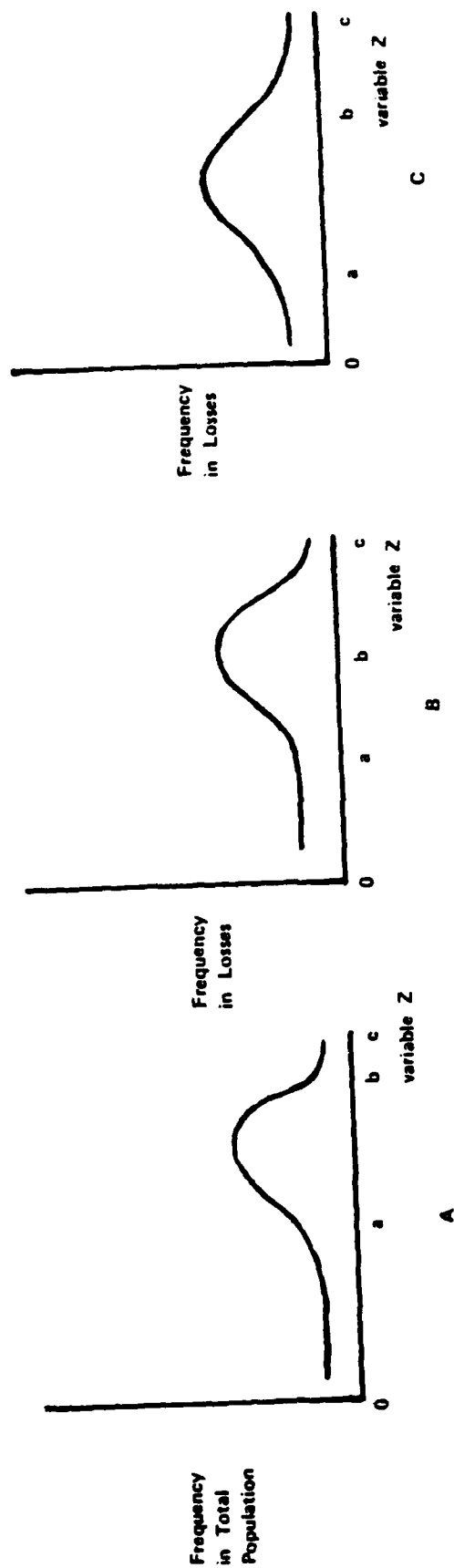
A and B are similar — no split on X
 A and C show disparate rates — split on variable X

Figure 5 - Dichotomies.



A and B are similar — no split on Y
A and C are dissimilar — split on Y
Possible categories are $\{0,2,4\}$, $\{1\}$, $\{3\}$

Figure 6 - Polytomies.



B may be similar to A, if segment (a,b) is too different, might use $\{(0,a), (b,c)\}$ and $\{(a,b)\}$ as separate categories

C — break Z into $\{(0,a), (a,b), \text{ and } (b,c)\}$ and use three separate categories.

Figure 7 - Discrete (or Pseudo-continuous).

variables. There is no general rule for the interval selection. However, the results of individual analyses should indicate the best choice for the given attribute.

Table 1 - Variable Types.

Variable	Type	Range/# of Categories
Transition	Polytomous	5
AFSC	Polytomous	13
Proficiency Pay	Discrete	0-3
Hazardous Duty Status	Discrete	0-100
Skill Level	Discrete	3-9 (coded 1-4)
Academic Level	Discrete	1-20
Method of Upgrade Training	Polytomous	7
Test Scores	Discrete	0-99 (coded 1-20)
Grade	Discrete	3-9
Year of Service	Discrete	0-21
Race	Polytomous	3
Sex	Dichotomous	2
Age	Discrete	0-∞
Marital Status	Polytomous	3
# of Dependents	Discrete	0-∞
Special Experience ID	Polytomous	17

The result of the histogram analysis is a list of driving attributes, the variables having a first order effect on transition rates. However, duplication of these effects among the variables may result because of a strong correlation between two or more variables. It is also conceivable that a second order interaction effect may be missed in the first pass through the variables. A refinement of the driving attribute identification process, as follows, should avoid these deficiencies.

1. Conduct first pass through the attributes as explained above.
2. From among the attributes falling out of the first pass, select the driving attribute that indicates the greatest disparity among transition rates.
3. Divide the whole population into subpopulations according to the relevant dichotomies, polytomies, or intervals for the most significant attribute.
4. Repeat the process for each of the individual subpopulations with a second attribute, third attribute, and so on selected for the variables remaining at each state of the stepwise process.

5. Terminate the process when none of the attributes remaining for each subpopulation fits the criteria for designation of driving attributes.

The output of this process is a sequenced list of driving attributes which contains all of the information necessary to cross-classify the airmen into homogeneous subpopulations.

In summary, categorical analysis involves the identification of attributes that coincide with nonuniform population transition rates. The search for driving attributes to define the subpopulations is conducted by comparing the frequency distributions of transitions.

THE TECHNIQUE OF DEPENDENCY ANALYSIS

A dependency analysis is conducted to determine the functional dependency of transition rates on the various attributes. It is directed toward devising a weighting scheme for the attribute relative to transition rates. Therefore, this type of analysis is based upon the investigation of interrelationships among the various attributes.

In dependency analysis, a transition is viewed as a response (dependent) variable and the Uniform Airman Record (UAR) file is considered a vector of explanatory (independent) variables. The dependency analysis technique was designed for the following hypothetical problem type.

A transition has been made. This transition is associated with information available in the 1975 UAR file for the transitioning airman. Airmen with sufficiently similar UARs have the same transition probabilities. The goal of dependency analysis is to develop the decision rule that best matches the actual results. That is, to build a model that most nearly would reproduce the observed transitions when the actual UAR records are used as input. Hence, the objective is to model the transition process with an algorithm, equation, or decision rule that will produce the expected transition file based on the input record.

Two forms of regression analysis were used as the dependency analysis technique. They are logit and discriminant analysis. A detailed description of logit analysis is provided in Appendix A. Discriminant analysis is detailed in Appendix B. Both appendix descriptions provide technical details and an analogy with Ordinary Least Square (OLS) regression.

The results of logit analysis may be applied to the PAM projection process in three ways. First, to identify variables that may contain significant information concerning transition probabilities. This provides not only a cross-check of the previous analysis of histograms, but also a valuable indication of which histograms are required.

The second use of the logit analysis is to assess the loss in projection accuracy which must occur when specific attributes are deleted from the cross-classification of data. Since there is a management cost for each variable included in the list of driving attributes, it is desirable to keep the list as short as possible. Logit analysis may be used to determine which attributes may be dropped from the list if a certain margin of projection error is allowable. For example, ignoring marital status may result in a two percent increase in classification error, but may present a four percent decrease in management costs. The logit analysis can aid in the assessment of cost/accuracy trade-offs.

The third application of the logit analysis is the impact assessment of Air Force personnel policy changes. This use of the logit analysis technique was discovered as a byproduct. It will be discussed later in a section of this report entitled "Modifying Transition Probabilities."

In summary, dependency analysis involves the investigation of interrelationships among various attributes. It employs a statistical regression technique to discover such interrelationships.

SUMMARY OF PROCEDURE ONE

Availability projections for subpopulations are highly similar to projections for homogeneous populations. The main differences are that each homogeneous subpopulations (within an AFSC) must be projected separately, and then the results for all subpopulations must be summed to obtain the AFSC results. In addition, a user should be aware of possible projection inaccuracies due to small size of homogeneous subpopulations.

Categorical and/or dependency analysis may be used in any given situation where the potential presence of driving attributes is under investigation. While categorical analysis tends to be easier to use, dependency analysis is more precise and powerful. Categorical analysis will tend to indicate more obvious driving attributes while dependency analysis will tend to indicate less obvious driving attributes.

It is preferable to apply the categorical analysis first. This will assure the user that all driving attributes have been discovered. Once discovered, it is relatively simple to proceed accurately with the projection process described in the following section.

PROCEDURE TWO: MAKING ACCURATE AVAILABILITY PROJECTIONS

Based on the results of categorical and/or dependency analysis in Procedure One, a user will be aware of any homogeneous subpopulations within a given AFSC (as indicated by the driving attributes). These homogeneous subpopulations should then be projected separately with the PAM to satisfy the assumptions of the Markov process. Each of the homogeneous subpopulation projections would follow the PAM application methodology using the existing PAM data base. The main differences between the projection procedures for entire homogeneous AFSCs (described in AFHRL-TR-79-66) and the projection procedures for homogeneous subpopulations within AFSCs are as follows.

1. Individual homogeneous subpopulations (within AFSCs) are extracted making use of the normal subpopulation selection features of the PAM and the driving attributes. The initial state matrices are filled according to the attributes selected as a result of the statistical analyses. The transition probability matrices are then computed as described in AFHRL-TR-79-66. (Note that changes in transition probabilities will be automatically made by the PAM depending upon which homogeneous subpopulations are being projected.)
2. The results of each homogeneous subpopulation projection (within a given AFSC) are summed at common time intervals to give the total HR availability projection for the entire AFSC.

Recall that Figure 1 presented the PAM projection process incorporating Procedures One and Two.

Considerations other than homogeneity also must be taken into account in the generation of subpopulations. Two of these considerations are computational cost and efficiency, both of which tend to constrain the number of subpopulations that can be used. As a result, accurate personnel availability projections are a function of the sample size and, thus, the number of subpopulations. When the total number of airmen in a transition state decreases, the accuracy of the estimated transition probabilities also decreases.

Since transition probabilities have a multinomial distribution, the error in the estimated transition probability is a function of the number of airmen in a particular state (see Appendix C for a derivation of this relationship). Hence, the user is faced with a trade-off. To accurately reflect a Markov process, as many subpopulations as necessary should be used to achieve homogeneity. To more accurately estimate the transition probabilities used in that Markov process, as few subpopulations as possible should be used.

The minimum value of n necessary for given confidence limits on the estimated transition probabilities are given in Appendix C. For example, a 90 percent confidence limit of $\pm .15$ on the estimated

transition probabilities requires that the state contain about 25 airmen, depending on the probability value. Operating on data which represent state sizes that are large enough ensures that spurious differences in transition rates are not identified as real differences.

DATA PROCESSING

Three statistical techniques were used in the analysis of transitions: histograms, logit analysis, and discriminant analysis. Using these techniques, a mathematical model was built that regressed the dependent variable on the independent variables. The results of the analyses are well suited to identify homogeneous subpopulations.

For these analyses, the abbreviated UAR was subjected to a recording process. Because computer programs available for these purposes require specific input data types and formats, it is necessary to convert UAR data to meet the requirements. The UAR data are primarily alphanumeric; the analyses operate only on numeric data. Therefore, the UAR data were altered as follows.

1. A variable was created to identify the actual transition of an airman in 1975.
2. The individual records were condensed so that data for the histogram analysis contained the attributes listed in Table 2 for 1975 and 1976.
3. All alphanumeric File Item Data Organizer (FIDO) codes were converted to numeric codes as shown in Table 2.

This alteration should be suitable for most computer programs that generate histograms. Packages for logit and discriminant analyses, however, probably require some polytomous variables to be expressed as a string of binary dummy variables. For example, the transition variable in the files created for the Statistical Analysis Subsystem (SASS) can be values from one to five in which:

- | | | |
|---|---|-----------|
| 1 | = | Loss |
| 2 | = | Transfer |
| 3 | = | Upgrade |
| 4 | = | Increment |
| 5 | = | Recruit |

The program used for logit and discriminant analyses required that the transition variable be expressed in the following steps.

1. Define 4 dummy variables--x1, x2, x3, and x4.
2. Set all variables to 0.
3. If airman i is lost, set x1 = 1.
4. If airman i transfers, set x2 = 1.
5. Continue in the above manner so that:
1000 = Loss
0100 = Transfer
0010 = Upgrade
0001 = Increment
0000 = Recruit

Table 2 - Data File.

Selected Attributes	Code Definitions of Attributes	New Code	
1. AFSC	325X0 = automatic flight control system	01	
	325X1 = instrument systems	02	
	328X0 = avionic communication	03	
	328X1 = avionic navigation systems	04	
	328X4 = inertial and radar navigation systems	05	
	423X0 = aircraft electrical systems	06	
	423X1 = environmental systems	07	
	423X3 = fuel systems	08	
	423X4 = pneudraulic system	09	
	429X2 = jet engines	10	
	431X1C = aircraft maintenance (jet, 1 or 2 engines)	11	
	431X1E = aircraft maintenance (jet, over 2 engines)	12	
	531X3 = airframe repair	13	
	External =	14	
2. Proficiency Pay	L = recruiting status with less than 6 month's experience	01	
	P = recruiting status, greater than 6 month's experience	01	
	Q = military training instructor in basic military training school	01	
	1 = lowest level of pro-pay	01	
	2 = second lowest level of pro-pay	02	
	3 = highest level of skill-rated pro-pay	03	
	4 = designated for pro-pay but not awarded	04	
	5 = pro-pay 2 status terminated	05	
	6 = pro-pay 3 status terminated	06	
	7 = 6 months as recruiter but below P-1 pay	07	
9 = assigned below group level P-1	09		
3. Hazardous Duty Status	2	personnel entitled to hazardous duty pay based on duty assigned	01
	3		
	4		
	5		
	6		
	7	= hostile pay only	01
	8		
	9		
	B	personnel entitled to hazardous duty pay + hostile duty pay	02
	C		
	D		
	X		
	1 = assigned to flying crew duty		03
A = assigned to flying crew and hostile duty		04	
Y = not assigned to hazardous duty		05	
Z = unknown		00	

Table 2 - Data File (continued)

Selected Attributes	Code Definitions of Attributes	New Code
4. Skill Level	1 = unskilled in AFSC	01
	3 = semi-skilled in AFSC	02
	5 = skilled in AFSC (technician)	03
	7 = advanced in AFSC (specialist)	04
	9 = superintendent in AFSC	05
5. Grade Level	31 = basic airman	01
	32 = airman	02
	33 = airman first class	03
	34 = sergeant	04
	35 = staff sergeant	05
	36 = technical sergeant	06
	37 = master sergeant	07
	38 = senior master sergeant	08
	39 = chief master sergeant	09
6. Year of Service (No. of years TAMS)	1975	01
	1974	02
	.	03
	.	04
	.	05
	19xx	75-xx
	1976	01
7. Race	C = caucasian	01
	N = black	02
	X = other	03
	Z = unknown	04
8. Marital Status	A = annuled	01
	D = divorced	02
	I = interlocutory	03
	L = legally separated	04
	W = widowed	05
	S = single	06
	M = married	07
9. Sex	M = male	01
	F = female	02

Table 2 - Data File (continued)

Selected Attributes	Code Definitions of Attributes	New Code
10. Test Scores (ASVAB)	administrative test electronics test general test mechanical test	% ranges from 0-99 all in increments of five
AFQT		
11. Highest Academic Level (years of school)	A = non-high school	03
	B } C } D } = high school completed (4 years) E }	04
	F } G } = high school + additional years I }	05
	H } I } = associate degree J }	06
	L = 3 years of college	07
	K } M } = 4 years of college N }	08
	O } 1 } = graduate work in progress	09
	P } 3 } = graduate work completed	10
	Q } E } = master's degree +	11
	R } S } = Doctoral degree	12
	T = second professional degree	16
	U = third professional degree	20
	Blank = no record of education	blank
12. Training Status Method of Meeting Upgrade	0 = not applicable	01
	2 = upgrade under new system—no board action	02
	3 = upgrade under new system—with board action	03
	4 = waived by classified in board	04
	5 = waived by HQ USAF	05
	6 = test deferred	06
	7 = passed test	07
	9 = unknown	08

Table 2 - Data File (continued).

Selected Attributes	Code Definitions of Attributes	New Code
13. Special Experience Identifier (1976)	501 = A-10	01
	522 = F/RF-4	02
	528 = F111D	03
	529 = F111F/FB-111	04
	532 = F-5	05
	533 = F-15	06
	535 = F-16	07
	538 = A7D	08
	545 = C-5	09
	546 = C9	10
	550 = C141	11
	569 = E-3A/B (AWACS)	12
	572 = C/AC/DC/HC/RC/WC-130	13 for 1976
	573 } = AC130, C130A/D/HC-130H/P/N	13 for 1975
	574 }	
	580 = C/KC/RC/EC-135/137	14
	586 = B-52/D/E/F in 1975, in 1976 is B-52	15
14. Age	1975 - year of birth	1975-year of birth
	.	.
	.	.
	.	.
	.	.
	xx - year of birth	xx-year of birth
15. Number of Dependents	Numeric	Numeric

Other attributes which were similarly expressed were race, marital status, sex, training status, and special experience identifier.

Several of the remaining polytomous variables can be transformed to simulate continuous variables. For example, the alphanumeric attribute, Highest Academic Level Achieved, was transformed to a variable, Years of School (as shown in Table 2). The decision to create dummy variables or to simulate continuous variables for each attribute must be based on the presence or absence of an underlying ordinal ranking for the individual attribute. A test for the existence of such a ranking is derived with the question: If two airmen have different FIDO codes for the same attribute, can it be said that one has more of that attribute than the other? If the answer is affirmative, there exists an ordinal ranking scheme. The problem then is to determine the basis of the ranking. In the case of race, for example, the designation is categorized without a quantitative implication. Thus, no ranking is possible. On the other hand, for the Highest Academic Level Achieved (FIDO AC-025) attribute, the designates do imply a continuum and the possibility of rank ordering. What remains is to select the basis, which is years of schooling in this case. Because K (Bachelor's degree) indicates a higher academic level than D (high school diploma), it is understood that more time was needed to achieve level K. Hence, years of schooling required to attain an academic level is a suitable basis for assigning pseudo-continuous values for this attribute.

It is recognized that this may introduce some error into the calculations. For example, some bachelor degrees require five years of study, while some individuals may legitimately document six, seven, or eight years of college without completing a degree. Since few enlisted personnel are likely to have advanced degrees, however, it is expected that this error would be very small.

MODIFYING TRANSITION PROBABILITIES

The PAM has presented a methodology for projecting the future availability of HR. However, the development of that methodology was based on the implicit assumption that the personnel transition process itself remains unchanged over the projective time period. In certain applications, this is a simplified and somewhat unrealistic representation of the Air Force personnel system. A more complete analysis requires that the following dynamic elements of that system be considered.

1. The variation of budgetary and authorization constraints.
2. The advent of new, more sophisticated weapon systems.
3. The conditions of the national labor market.

Each of these elements and others directly affect the quantity and quality of available manpower.

A need, therefore, exists to assess the likely impact of these dynamic elements on personnel availability, especially when questions arise concerning the likely effects of a policy change. Presently, the PAM system can generate at least qualitative answers to such questions. By using the results of a logit analysis (detailed in Appendix A), it was found that a PAM user may systematically examine the quantitative effect of policy changes on future HR availability. However, the user must know the effect of each policy change on the transition probabilities.

A logit equation gives the transition probability of an airman with a specific set of attributes. It is easy to compute a new transition probability if the set of attributes is perturbed. Thus, the user may assess the impact of changing any attribute or combination of attributes on the transition probabilities. Given sufficient familiarity with the equation, it is possible to determine which variables significantly influence the transition probabilities and to determine the degree of influence.

Once the influences are determined, changes in the transition probabilities can be made using the existing PAM maintenance function. This function allows the user to modify manually any of the elements of the transition probability matrices. The link between perturbed attributes and resultant changes in transition probabilities then can be used to compute a new HR availability projection. A user who has prior knowledge of the effect of a policy change in transition probabilities may directly change the transition matrices and generate new HR availability projections. If, instead, the user has prior knowledge of the effect a change may have on the individual attributes, the logit equations will give an indication of how the effects may be translated into variations of the transition probabilities.

The ability of the logit equations to estimate the sensitivity of the transition probabilities to individual attributes can also be used to identify candidate policy change options. Assuming that there is a

particular transition rate that the PAM user desired to change, the logit functions will be useful in determining which attribute changes are most likely to result in that desired rate change. Again, the user must have some prior knowledge of the ways in which policy changes can affect the incidence of personnel attributes within the population and/or transition rates for that population.

In summary, future HR availability will be affected by expected or proposed changes in Air Force personnel policy. Such changes should be evaluated through projections of future HR availability. This evaluation can now be accomplished employing the technique of logit analysis along with built-in PAM program features, when the user has prior knowledge of which policy changes might be implemented in the future.

EXTENDING THE RANGE OF THE PAM

The PAM was initially designed to project the future availability of Air Force technical personnel. The projective capability was to be applied only to technical personnel because they were considered to be composed of homogeneous populations. This belief was developed because technical personnel must successfully pass through a number of screening devices during selection and training. Given this assumed homogeneity, their future availability could easily be computed using a fixed mathematical model.

The effort described in this report was an attempt to check the validity of the mathematical projection approach, and thus, the assumption of homogeneity. As described earlier, procedures were developed to discover any personnel attributes which might upset the homogeneity assumption and, thus, invalidate the use of a final mathematical model (particularly, the Markov process model which is the basis of the PAM).

Contrary to the assumption made for technical personnel, non-technical personnel were not assumed to be likely candidates for application of the PAM because, for a variety of reasons, they were not considered likely to be homogeneous.

Considering the results of the effort reported within this report, it now seems likely that the PAM can be accurately applied to non-technical personnel. The statistical techniques of categorical and dependency analysis can be applied successfully to populations of non-technical personnel to discover the driving attributes which affect their transitions. Once these driving attributes are identified, homogeneous subpopulations at a level below that of AFSC can also be identified and used to project their future availability. Summing the projections for the homogeneous subpopulations for a given year will result in the availability of the entire population within the AFSC for the specified year.

The transition probabilities for homogeneous subpopulations of nontechnical personnel may also be varied based upon expected or proposed changes in Air Force personnel policy. This is accomplished using the technique of logit analysis. As long as homogeneity of subpopulations is present, such modified projections should be accomplished as easily as those for technical personnel.

In addition to achieving the objective of the effort, the result of the effort described in this report is a dramatic widening in the range of potential applications for the PAM. Along with technical personnel, it is now entirely possible to project the future availability of all Air Force personnel, regardless of AFSC. The PAM methodology thus becomes a far more powerful planning tool.

CONCLUSION

The work described in this report was designed to ascertain the validity of the Markov process as it relates to Air Force personnel possessing technical AFSCs. To achieve this objective two procedures were developed. Procedure One is composed of two parallel statistical techniques (categorical analysis and dependency analysis) and was developed for application to the PAM data base. The purpose of this procedure was to distinguish those driving attributes which may impact personnel transition rates.

Procedure Two was developed to modify the projection process in the event that impacting attributes were found in the data base. The procedure permits homogeneous subpopulations to be formed and then summed to provide a total availability for a given AFSC.

In AFHRL-TR-79-68, Procedure One is applied directly to the existing PAM data base. The results indicate that there are no impacting personnel attributes other than the technical AFSC; therefore, the application of Procedure Two was not required. However, the objective of this effort was achieved, and the validity of the Markov model was supported as valid for projecting the future availability of Air Force personnel possessing technical AFSCs.

In addition, the two procedures have resulted in a general methodology which can be applied to any population regardless of AFSC. Use of the methodology permits a user to recognize those personnel attributes which affect transition rates and thus identify related homogeneous subpopulations. Once the homogeneous subpopulations have been specified, the projection procedure outlined in this report may be employed to make accurate availability projections. Thus, this general methodology permits the Markov model to be used with populations which are not initially homogeneous. As a result, the range of application of the PAM can be extended to all Air Force personnel.

A further byproduct of this effort involves the technique of logit analysis. When reviewing this technique as related to dependency analysis, it was found that logit analysis has application in reflecting expected or proposed changes in Air Force personnel policy. As a result, it is now possible to accurately modify PAM projections based upon such changes.

REFERENCES

- Goclowski, J. C., LoFaso, A. J., Peskoe, S. E., & Baran, H. A. Air Force personnel availability analysis: A description of the Personnel Availability Model (PAM). AFHRL-TR-79-66. Wright-Patterson AFB, OH: Advanced Systems Division, Air Force Human Resources Laboratory, 1980.
- Goclowski, J. C., LoFaso, A. J., Peskoe, S. E., & Baran, H. A. Air Force personnel availability analysis: Program description for the Personnel Availability Model (PAM). AFHRL-TR-79-68. Wright-Patterson AFB, OH: Advanced Systems Division, Air Force Human Resources Laboratory, 1980.

APPENDIX A - LOGIT ANALYSIS

The basic concept of logit analysis [1,2] is similar to Ordinary Least Squares (OLS) multiple regression techniques. If x_i is a vector of independent variables and y_i is the dependent variable, then OLS determines the linear function of x_i that best approximates the probability distribution of $E(y_i|x_i)$, that is, the expected value of y_i given x_i . A vector of coefficients, β , is sought so that $x_i' \beta$ is the best estimate for $E(y_i|x_i)$. The criterion for "best" is to choose β such that $E[(y_i - x_i' \beta)^2]$ is minimized, producing the least squares estimate.

In this research, the objective is to determine the functional relationship between transitions and attributes, where x_i represents the abbreviated UAR file for airman i and y_i represents the transition an airman made in 1975. Considering losses specifically, $y_i = 1$ if the airman left the Air Force in 1975; otherwise, $y_i = 0$.

If all the variables were continuous, one of several software packages that perform the OLS multiple regression could be readily applied. However, the variables are not continuous. Several variables are dichotomous or polytomous, and the others are discrete. Of most importance here is the fact that the dependent variable is polytomous. That is, an airman upgrades, leaves, transfers, or increments. He cannot partially upgrade or half leave. (Refer to Table 1 earlier in this report for a list of variables and their respective types--dichotomous, polytomous, or discrete.)

A problem inherent in the structure of OLS regression is that the estimates of the dependent variable y_i (that is, the $x_i' \beta$) can take on any numerical value despite the fact that $y_i = \{0, 1\}$ only. Therefore, OLS estimators of $x_i' \beta$ can be unreasonable. Restricted least squares in which β is estimated by OLS with the constraint that $0 \leq x_i' \beta \leq 1$, is frequently a viable alternative. However, this is not applicable to the problem at hand due to the dependence of the variance of y on i which results in inefficient estimates of β . More observations are required to achieve the same confidence limits on β than would be the case if a procedure were used that estimates independent of y 's dependence on i . Although the reformulation of the problem in logit analysis obviates the difficulties expressed above, there is some generality and lack of precision.

Discrete dependent variable regression, of which logit analysis is one technique, had its origin in the science of bioassay. In this formulation, as above, y_i is the i th dependent variable and equals 1 if airman i left the Air Force. It is assumed that behind this dichotomous variable (y) is an unobservable variable (z) with a continuous distribution.

Furthermore, each airman has a particular threshold level (z_i), which is dependent upon x_i , such that $y_i = 1$ if $Z_i < x_i' \beta$ and $y_i = 0$ if $z_i \geq x_i' \beta$. As before, $x_i' \beta$ is a linear function of x in which x_i denotes the attributes and is the vector of coefficients to be estimated. The algebraic simplicity and computational tractability of the logistic distribution and the fact that many natural populations follow an approximate logistic function make the choice for the distribution of z quite reasonable. The necessity of the assumption of a probabilistic functional form between y and z is the loss of generality previously noted.

If $F(\cdot)$ denotes the cumulative density function of z given x , the following equations are true.

$$\begin{aligned} P \{ y_i = 1 \} &= P \{ z_i \leq x_i' \beta | x_i \} = F(x_i' \beta) \\ P \{ y_i = 0 \} &= 1 - F(x_i' \beta) \end{aligned}$$

The best estimate of β in this problem is chosen to be the Maximum Likelihood Estimate (MLE). Likelihood, a basic concept of mathematical statistics, can be thought of as a measure of the ability of a probability density function to explain the variable y . It is the joint probability of the y_i and is denoted $L(y; f, \beta)$ where f is a given density function. Here, the density function is the logistic, and β is a parameter of the distribution. If β_1 and β_2 are two estimates for β , then β_1 is a better estimate of β if $L(y; f, \beta_1) > L(y; f, \beta_2)$. We seek the estimate of β that gives the maximum value of $L(y; f, \beta)$. In the formulation of OLS, the least squares estimate is the MLE for where $f(\cdot)$ is the normal distribution. The likelihood function is given in the following equation.

$$L(y) = \prod_{i=1}^n F(x_i' \beta)^{y_i} [1 - F(x_i' \beta)]^{1-y_i}$$

If $F(t)$ represents the logistic function, $F(t) = (1 + \exp(-t))^{-1}$, $-\infty < t < \infty$, $L(y)$ becomes:

$$\begin{aligned} L(y) &= \prod_{i=1}^n \left[\frac{1}{1 + \exp(-x_i' \beta)} \right]^{y_i} \left[1 - \frac{1}{1 + \exp(-x_i' \beta)} \right]^{1-y_i} \\ &= \frac{\exp \beta' \sum_i x_i y_i}{\prod_i [1 + \exp(x_i' \beta)]} \end{aligned}$$

Note that $t^* = \sum_i x_i y_i$ is sufficient for β , where a sufficient statistic is one that summarizes all of the relevant information about β supplied by the sample data.

The MLE of β , $\hat{\beta}$, is found by setting the derivative of $\log L(y)$ equal to zero and solving for β . Thus, the following equations are derived.

$$L^* = \log L = \beta' t^* - \sum_i \log (1 + \exp(x_i' \beta)) = 0$$

$$t^* = \sum_i [1 + \exp(-x_i' \beta)]^{-1} x_i = \sum_i x_i y_i$$

The equation is solved numerically for $\hat{\beta}$ in the logit analysis program. An additional and useful feature of this program is the performance of the regression in a stepwise manner. This means that the attributes are entered into the linear equation $x' \beta$ one at a time. The order of entry is again based on a maximum likelihood criterion. If x^1 , x^2 , and x^3 are three attributes, and $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ are the respective MLEs when each x_j is used alone, then x^1 will be chosen first only if $L(y; f(x^1), \hat{\beta}_1)$ is greater than $L(y; f(x^2), \hat{\beta}_2)$ and $L(y; f(x^3), \hat{\beta}_3)$. If x^1 is entered first, x^2 will be entered only if $L(y; f(x^1, x^2), \hat{\beta}_{12}) > L(y; f(x^1, x^3), \hat{\beta}_{13})$. The program will search through all possible attributes x_j , building a linear function of x s and $\hat{\beta}$ until either all attributes are entered or the addition of any remaining attribute will not significantly increase the likelihood of the whole model. Thus, if x^1 , x^2 , and x^3 exist, and x^1 , x^2 are already entered, then x^3 will be entered only if:

$$\log L(y; f(x^1, x^2, x^3), \hat{\beta}_{123}) - \log L(y; f(x^1, x^2), \hat{\beta}_{12}) > \Sigma$$

In this case, Σ is a user-specified value that sets the size or significant level of the test. It is claimed that $\log L(y)$ is distributed approximately as the χ^2 with 1 degree of freedom so that $\Sigma = 3.84$ sets the significance level at 5 percent. In other words, no attributes are used unless they carry sufficient information about transitions to significantly increase the ability to explain y .

Logit analysis determines the linear function of the attributes that best explains the transitions using as few attributes as necessary to reach a given confidence level.

APPENDIX B - DISCRIMINANT ANALYSIS

Considering Air Force losses and retentions specifically, the logit analysis gives the probability that a specific airman will leave the Air Force and the discriminant analysis provides a measure of how similar he is to each of the two groups. Logit analysis estimates loss probability, conditioned on the attributes. Discriminant analysis [2,3] classifies according to attributes conditioned on the transitions.

If there are n attributes under consideration, X_i is an n -dimensional vector that identifies an airman's location in the attribute space. In the discriminant analysis formulation, $X'B_1$ and $X'B_2$ are two linear functions of the X s that are determined for maximizing the square distance between group centroids in the attribute space. Let the sub-population under investigation be partitioned into two groups:

- Group 1 The set of airmen that remained in the Air Force.
- Group 2 The set of losses.

The goal of discriminant analysis is to find the two linear functions of X (F_1 and F_2) that satisfy the following conditions.

1. If $X_i(1)$ is the UAR of airman i in Group 1 and $X_j(2)$ is the UAR of airman j in Group 2;
2. And if $Y_i(1) \equiv F_1(X_i(1))$ and if $Y_j(2) \equiv F_2(X_j(2))$

then the average over i of $Y_i(1)$ and the average over j of $Y_j(2)$ (these averages are the group centroids) are at a maximal distance from each other in the n -dimensional attribute space.

The functions F_1 and F_2 are approximated by $X_i' \beta_1$ and $X_i' \beta_2$, respectively. Discriminant analysis serves to determine the β_1 and β_2 values that meet the listed criteria.

This is an analysis in which an attempt is made to place a given airman into Group 1 or Group 2, based on his individual set of attributes. It is possible to compute the posterior probability that the airman is in either group. The solution of the classification problem is to assign the airman to the group for which the posterior probability of membership is largest. In effect, the question that has been answered is: In terms of transitioning, to which group of airmen is the individual airman most similar?

As with logit analysis, the discriminant analysis may be performed in a stepwise manner by choosing, at each stage of an iterative process, the independent variable that contains the most information about the transitions. However, although this technique may be used in much the same manner as the logit analysis, subtle differences in the interpretation of results must be recognized.

REFERENCES TO APPENDICES A AND B

1. Nerlove, Marc and Press, S. James Univariate and Multivariate Log-Linear and Logistic Models. Santa Monica, CA: Rand Corporation, December 1973, Pub. No. R-1306-EDA/NIH, p. 134.
2. Gnanadesikan, R. Methods for Statistical Data Analysis of Multivariate Observations. New York: Wiley and Sons, 1977, p. 311.
3. Cooley, William W. and Lohnes, Paul R. Multivariate Data Analysis. New York: John Wiley and Sons, 1971, p. 364.

PRECEDING PAGE BLANK-NOT FILMED

APPENDIX C - CONFIDENCE LIMITS

The transition probabilities used by the PAM are computed with the maximum likelihood estimator appropriate to a Markov process with stationary transition probabilities. Consider a Markov state containing airmen and a given type of transition out of the state that occurs with probability p . The number of airmen who actually make that transition is distributed as a binomial random variable (expressed as Equation C1).

$$\text{Equation C1} \quad P \{ X = x \} = \left[\frac{n!}{(n-x)! x!} \right] p^x (1-p)^{n-x}$$

The operands are:

- X A random variable that counts the number of airmen making the transition.
- x A specific value of X .
- $P \{ \cdot \}$ The probability of the event within the braces.
- $n!$ The factorial operation $(n) (n-1)(n-2) \dots (1)$.

In general, p is unknown and must be estimated by observing the system and counting the number of airmen who made specified transitions. The maximum likelihood estimator for p is expressed as:

$$p = \frac{x}{n}$$

The operands are:

- x The number of airmen observed to make a given transition from a given state.
- n The population of the state.

The crucial question is: Given a data sample (basically x and n), how far off is the estimated value \hat{p} from the true value, p ? This question is more precisely addressed in Equation C2 which states that the probability that \hat{p} (the estimated value) deviates from p (the true value) by an amount greater than or equal to ϵ is less than or equal to α . Thus, the smaller that ϵ and α are, the more confidence that can be placed on the accuracy of \hat{p} .

$$\text{Equation C2} \quad P \{ |\hat{p} - p| \geq \epsilon \} \leq \alpha$$

The operands are:

- $|\cdot|$ The absolute value of the argument within the vertical lines.
- ϵ The magnitude of the error in the estimate $0 < \epsilon < 1$
- α The upper limit on the probability a magnitude ϵ .

Equations C1 and C2 can be used to analyze the small numbers problem mentioned in this report. For state populations less than 25, there is a large probability of a significant error in the estimated transition probability. It follows then, that Equation C2 should be evaluated to determine the population size needed to estimate p with a required accuracy. That is, given a required ϵ and a , what is the smallest acceptable sample size (Markov state population)?

The minimum sample size is derived from Equation C2. Substituting $\frac{x}{n}$ for \hat{p} , the equation becomes:

$$P \left\{ \left| \frac{x}{n} - p \right| \geq \epsilon \right\} \leq a$$

Removing the absolute value expression, the expression is written as:

$$P \left\{ x < n(p - \epsilon) \right\} + P \left\{ x \geq n(p + \epsilon) \right\} \leq a$$

Using the fact that x is an integer, the equation then becomes:

$$P \left\{ x \geq [n(p + \epsilon)] + 1 \right\} + P \left\{ x \leq [n(p - \epsilon)] \right\} \leq a$$

Using Equation C1, the expression can be explicitly evaluated (as Equation C3) for a given sample size n .

Equation C3

$$\sum_{x=[n(p+\epsilon)]+1}^n \binom{n}{x} p^x (1-p)^{n-x} + \sum_{x=0}^{[n(p-\epsilon)]} \binom{n}{x} p^x (1-p)^{n-x} \leq a$$

$$\text{where } \binom{n}{x} = \frac{n!}{(n-x)! x!}$$

For a given a and ϵ , Equation C3 is evaluated at successive values of n ($n=0, 1, 2, \dots$) until the sums of the left-hand side are less than a . Note that ϵ enters the expression through the summation indices, and that the sample size also depends on the probability (p) which is being estimated. (For a larger n , the normal approximation to the binomial can be used to evaluate Equation C3.)

Using the above analytical procedures, a table was produced consisting of sample sizes (n) as a function of the transition probability (p), the error (ϵ) of the estimated \hat{p} , and the probability of error (α) of magnitude ϵ . This list is contained in Table C1. A researcher using the PAM should consult this appendix in order to determine the probable accuracy of the transition probability matrices computed for a given subpopulation. For example, if a researcher required that the estimated transition probability be within .15 of the true value with a confidence limit of $.9 = (1 - \alpha)$, the state population would have to be approximately 25, depending on the value of the transition probability (refer to Table C1).

Table C1 - Confidence Limits.

Error in Transition Probability	Confidence Limit	Transition	State Population
0.050	0.950	0.100	134
0.050	0.950	0.200	232
0.050	0.950	0.300	315
0.050	0.950	0.400	365
0.050	0.950	0.500	371
0.050	0.950	0.600	365
0.050	0.950	0.700	315
0.050	0.950	0.800	233
0.050	0.950	0.900	134
0.050	0.900	0.100	94
0.050	0.900	0.200	164
0.050	0.900	0.300	223
0.050	0.900	0.400	254
0.050	0.900	0.500	262
0.050	0.900	0.600	254
0.050	0.900	0.700	223
0.050	0.900	0.800	165
0.050	0.900	0.900	94
0.050	0.850	0.100	74
0.050	0.850	0.200	124
0.050	0.850	0.300	163
0.050	0.850	0.400	194
0.050	0.850	0.500	191
0.050	0.850	0.600	194
0.050	0.850	0.700	163
0.050	0.850	0.800	125
0.050	0.850	0.900	74
0.050	0.800	0.100	54
0.050	0.800	0.200	92
0.050	0.800	0.300	135
0.050	0.800	0.400	154
0.050	0.800	0.500	151
0.050	0.800	0.600	154
0.050	0.800	0.700	135
0.050	0.800	0.800	93
0.050	0.800	0.900	54
0.050	0.750	0.100	47
0.050	0.750	0.200	72
0.050	0.750	0.300	103
0.050	0.750	0.400	125
0.050	0.750	0.500	122
0.050	0.750	0.600	125
0.050	0.750	0.700	103
0.050	0.750	0.800	73
0.050	0.750	0.900	47

Table C1 - Confidence Limits (continued).

Error in Transition Probability	Confidence Limit	Transition	State Population
0.100	0.950	0.100	35
0.100	0.950	0.200	57
0.100	0.950	0.300	78
0.100	0.950	0.400	86
0.100	0.950	0.500	92
0.100	0.950	0.600	93
0.100	0.950	0.700	78
0.100	0.950	0.800	57
0.100	0.950	0.900	36
0.100	0.900	0.100	30
0.100	0.900	0.200	37
0.100	0.900	0.300	53
0.100	0.900	0.400	62
0.100	0.900	0.500	62
0.100	0.900	0.600	63
0.100	0.900	0.700	53
0.100	0.900	0.800	37
0.100	0.900	0.900	31
0.100	0.850	0.100	25
0.100	0.850	0.200	27
0.100	0.850	0.300	43
0.100	0.850	0.400	46
0.100	0.850	0.500	47
0.100	0.850	0.600	49
0.100	0.850	0.700	43
0.100	0.850	0.800	27
0.100	0.850	0.900	26
0.100	0.800	0.100	20
0.100	0.800	0.200	27
0.100	0.800	0.300	33
0.100	0.800	0.400	36
0.100	0.800	0.500	37
0.100	0.800	0.600	39
0.100	0.800	0.700	33
0.100	0.800	0.800	27
0.100	0.800	0.900	21
0.100	0.750	0.100	17
0.100	0.750	0.200	17
0.100	0.750	0.300	28
0.100	0.750	0.400	26
0.100	0.750	0.500	27
0.100	0.750	0.600	33
0.100	0.750	0.700	28
0.100	0.750	0.800	17
0.100	0.750	0.900	17

Table C1 - Confidence Limits (continued).

Error in Transition Probability	Confidence Limit	Transition	State Population
0.150	0.950	0.100	31
0.150	0.950	0.200	20
0.150	0.950	0.300	36
0.150	0.950	0.400	39
0.150	0.950	0.500	37
0.150	0.950	0.600	39
0.150	0.950	0.700	36
0.150	0.950	0.800	20
0.150	0.950	0.900	31
0.150	0.900	0.100	24
0.150	0.900	0.200	15
0.150	0.900	0.300	25
0.150	0.900	0.400	24
0.150	0.900	0.500	28
0.150	0.900	0.600	24
0.150	0.900	0.700	20
0.150	0.900	0.800	15
0.150	0.900	0.900	25
0.150	0.850	0.100	20
0.150	0.850	0.200	12
0.150	0.850	0.300	18
0.150	0.850	0.400	20
0.150	0.850	0.500	17
0.150	0.850	0.600	22
0.150	0.850	0.700	18
0.150	0.850	0.800	12
0.150	0.850	0.900	21
0.150	0.800	0.100	17
0.150	0.800	0.200	12
0.150	0.800	0.300	16
0.150	0.800	0.400	15
0.150	0.800	0.500	14
0.150	0.800	0.600	15
0.150	0.800	0.700	16
0.150	0.800	0.800	12
0.150	0.800	0.900	17
0.150	0.750	0.100	16
0.150	0.750	0.200	9
0.150	0.750	0.300	12
0.150	0.750	0.400	11
0.150	0.750	0.500	11
0.150	0.750	0.600	11
0.150	0.750	0.700	12
0.150	0.750	0.800	9
0.150	0.750	0.900	17

Table C1 - Confidence Limits (continued).

Error in Transition Probability	Confidence Limit	Transition	State Population
0.200	0.950	0.100	29
0.200	0.950	0.200	18
0.200	0.950	0.300	18
0.200	0.950	0.400	22
0.200	0.950	0.500	23
0.200	0.950	0.600	22
0.200	0.950	0.700	18
0.200	0.950	0.800	18
0.200	0.950	0.900	29
0.200	0.900	0.100	23
0.200	0.900	0.200	13
0.200	0.900	0.300	14
0.200	0.900	0.400	14
0.200	0.900	0.500	13
0.200	0.900	0.600	14
0.200	0.900	0.700	10
0.200	0.900	0.800	13
0.200	0.900	0.900	23
0.200	0.850	0.100	19
0.200	0.850	0.200	10
0.200	0.850	0.300	8
0.200	0.850	0.400	12
0.200	0.850	0.500	12
0.200	0.850	0.600	12
0.200	0.850	0.700	8
0.200	0.850	0.800	11
0.200	0.850	0.900	19
0.200	0.800	0.100	17
0.200	0.800	0.200	10
0.200	0.800	0.300	6
0.200	0.800	0.400	9
0.200	0.800	0.500	9
0.200	0.800	0.600	9
0.200	0.800	0.700	6
0.200	0.800	0.800	11
0.200	0.800	0.900	17
0.200	0.750	0.100	14
0.200	0.750	0.200	8
0.200	0.750	0.300	6
0.200	0.750	0.400	9
0.200	0.750	0.500	3
0.200	0.750	0.600	9
0.200	0.750	0.700	6
0.200	0.750	0.800	8
0.200	0.750	0.900	14

END
DATE
FILMED
10-80
ETIC





AD-AD88 801 DYNAMICS RESEARCH CORP WILMINGTON MASS P/G 6/9
AIR FORCE PERSONNEL AVAILABILITY ANALYSIS; APPLICATION TECHNIQUE--ETC;
AUG 80 J C GOCLOWSKI, S E PESKOE, A J LOFARO F33615-77-C-0032
UNCLASSIFIED AFMRL-TR-79-67 NL

2 OF 2

AD 2
000000



AD-AD88 801
AD 2
000000

AD-AD88 801
AD 2
000000

END
DATE
FORMED
6-81
DTIC

AD
08

SUPPLEMENTARY

INFORMATION

DEPARTMENT OF THE AIR FORCE
AIR FORCE HUMAN RESOURCES LABORATORY (AFSC)
BROOKS AIR FORCE BASE, TEXAS 78235



REPLY TO
ATTN OF: TSR

Errata

16 JAN 1981

SUBJECT: Removal of Export Control Statement

TO: Defense Technical Information Center
Attn: DTIC/DDA (Mrs Crumbacker)
Cameron Station
Alexandria VA 22314

1. Please remove the Export Control Statement which erroneously appears on the Notice Page of the reports listed ~~AD-A088801~~. This statement is intended for application to Statement B reports only.

2. Please direct any questions to AFHRL/TSR, AUTOVON 240-3877.

FOR THE COMMANDER

Wendell L Anderson

WENDELL L. ANDERSON, Lt Col, USAF
Chief, Technical Services Division

1 Atch
List of Reports

Cy to: AFHRL/TSE

AD-A088801